Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Virus Research 285 (2020) 198022

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

COVID-19: CADD to the rescue

Abdulmujeeb T. Onawole[a], Kazeem O. Sulaiman[b,*], Temitope U. Kolapo[c,d], Fatimo O. Akinde[e], Rukayat O. Adegoke[f]

[a] Department of Chemistry, King Fahd University of Petroleum and Minerals, Dhahran, 31261 Saudi Arabia
[b] Department of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan S7N 5C9, Canada
[c] Department of Veterinary Parasitology and Entomology, University of Ilorin,P.M.B. 1515, Ilorin, Nigeria
[d] Department of Veterinary Microbiology, University of Saskatchewan, 52 Campus Drive, Saskatoon, Saskatchewan S7N 5B4, Canada
[e] School of Pharmacy, Shenyang Pharmaceutical University, Shenyang 110016, Liaoning, China
[f] Department of Pure and Applied Biology, Ladoke Akintola University of Technology, P.M.B. 4000, Ogbomoso, Nigeria

ARTICLE INFO

Keywords:
COVID-19
Coronavirus
CADD
Zoonotic diseases
SARS-CoV-2
Virtual screening

ABSTRACT

The recent outbreak of the deadly COVID-19 disease, being caused by the novel coronavirus (SARS-CoV-2), has put the world on red alert as it keeps spreading and recording more fatalities. Research efforts are being carried out to curtail the disease from spreading as it has been declared as of global health emergency. Hence, there is an exigent need to identify and design drugs that are capable of curing the infection and hinder its continual spread across the globe. Herein, a computer-aided drug design tool known as the virtual screening method was used to screen a database of 44 million compounds to find compounds that have the potential to inhibit the surface glycoprotein responsible for virus entry and binding. The consensus scoring approach selected three compounds with promising physicochemical properties and favorable molecular interactions with the target protein. These selected compounds can undergo lead optimization to be further developed as drugs that can be used in treating the COVID-19 disease.

1. Introduction

The reemergence of the coronavirus has taken the world by storm such that on the 30th of January 2020, the World Health Organization (WHO) declared it a public health emergency and later in February 2020, WHO officially named the novel coronavirus disease as COVID-19 (World Health Organization, 2020). The SARS-CoV-2 is the fourth zoonotic coronavirus to emerge in the last twenty years. The first two, the Severe Acute Respiratory Syndrome (SARS-CoV) and the Middle East Respiratory Syndrome (MERS-CoV), appeared in 2002 (Zhong et al., 2003) and 2012 (Sousou, 2015) respectively while in 2017, the Swine Acute Diarrhea Syndrome (SADS-CoV) affected the swine livestock (Cui et al., 2019). These diseases are known to be zoonotic and being transmitted by bats (Drexler et al., 2014) and it is suggested that the novel coronavirus, that was first identified in Wuhan, China, (SARS-CoV-2) is not an exception as some researchers had earlier predicted that there might be another zoonotic coronavirus outbreak in early 2019 (Fan et al., 2019).

Like other similar virus, recent updates show that the SARS-CoV-2 now spread from man to man, although it is presumed to be zoonotic in origin. While the genetic research confirmed SARS-CoV-2 is originated in bats, there are other speculations that other wild animals could serve as intermediary between bats and man, with pangolins leading as primary suspects as the intermediary in the case of Wuhan coronavirus (Liu et al., 2019). SARS-CoV-2 spread like other cold viruses and early symptoms include, but not limited to, runny nose, severe cough, sore throat, difficulty in breathing, etc. With several precautions and awareness on the deadly virus and its spread, more cases of infection are being anticipated globally. The known cases are being managed with supplemental oxygen and conservative fluid administration as there is currently no approved vaccine or antiviral agent to treat the infection by SARS-CoV-2 and the concerned researchers are continuously working on developing a vaccine or drug for novel coronavirus (Li et al., 2020; Liu et al., 2019).

With the continuous rise in number of confirmed cases since the outbreak of the SARS-CoV-2, a fast and reliable tool such as computer-aided drug design (CADD) is of the essence. CADD is a renowned tool in the pharmaceutical industry and it does not only save time but also helps to cut costs of designing drugs. Virtual screening (VS) is one of the methods used in CADD and it enables screening of many compounds in a relatively short time compared to the high throughput screening via laboratory experiments (Kapetanovic, 2008; Leelananda and Lindert, 2016; Macalino et al., 2015; Manas and Green, 2017; Melo-Filho et al., 2019). Moreover, molecular docking as well as machine learning, can be used in virtual screening and these further enable effectiveness of VS

* Corresponding author.
E-mail address: kosulaiman2008@yahoo.com (K.O. Sulaiman).

https://doi.org/10.1016/j.virusres.2020.198022
Received 12 February 2020; Received in revised form 30 April 2020; Accepted 11 May 2020
Available online 15 May 2020
0168-1702/

(Mori et al., 2012; Pereira et al., 2020). For example, VS was employed in the development of approved drugs such as Aggrastat, a fibrinogen receptor, and Cevoglitazar, an effective PPAR-α/γ dual agonist for diabetes treatment (Clark, 2008). Recently, the use of consensus scoring has been acknowledged to also improve the enrichment of true positives and improve hit rates (Charifson et al., 1999; Clark et al., 2002; Feher, 2006). Herein, we employed computer-aided drug design that entails the use of consensus scoring to combine both molecular docking and machine learning VS method to discover potential inhibitors of the surface glycoprotein of the SARS-CoV-2 which is responsible for virus binding and entry. Our results identify three compounds with promising physicochemical properties and favorable molecular interactions with the target protein, and by extension, these identified compounds can undergo lead optimization to be further developed as drugs that can be used in treating the COVID-19.

2. Methodology

2.1. Target protein preparation

Due to the recent emergent of the coronavirus disease, there are not many crystal structures of high resolution of the virus. As at the time the bulk of this work was completed, there was no crystal structure of the spike protein. Hence, homology modeling (Haddad et al., 2020; Krieger et al., 2003; Xiang, 2006) was employed. However, as at the end of March 2020, there are currently about 100 crystal structures of the SARS-CoV-2 deposited in the protein databank (www.rscb.org, Burley et al., 2018). It is intended that the best among these structures, particularly the spike proteins (PDB ID: 6VSB, 6VYB, 6LZG, 6MOJ and 6VXX) (Lan et al., 2020; Walls et al., 2020; Wrapp et al., 2020), would be used for future works. The surface spike glycoprotein functions as the cell-attachment recognition site hence, the reason for its consideration as the target protein as this is the critical part of the virus responsible for virus entry and binding (Gralinski and Menachery, 2020). The target protein was prepared using homology modeling with the aid of the Raptor program (Peng and Xu, 2011). The template protein used for the homology model is PDB ID: 5 × 58 (Yuan et al., 2017). The sequence of the glycoprotein utilized to build the homology model was retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (Wu et al., 2020) with the accession number: MN908947. The modeled structure was validated using the Ramachandran plot analysis with the aid of RAMPAGE webtool (Lovell et al., 2003; Ramachandran and Sasisekharan, 1968) and PrankWeb server (Jendele et al., 2019) to validate the amino residued involved in the protein-ligand interactions. The ConSurf web program (Ashkenazy et al., 2016) was used for the multiple sequence alignment (MSA) analysis, conserved score and phylogentic analysis.

2.2. Virtual screening

The MCULE full database (Kiss et al., 2012) with exactly 44,704,142 compounds, as at that the time of this work, was used for the first virtual screening experiment. A blind docking (Grosdidier et al., 2009) was carried out, which covered the whole of the protein since no binding pocket has been determined from experiments yet with the following parameters -1.298, -7.617 and 191.965 for X, Y and Z axes respectively. These coordinates represent the binding site area. The MCULE database was filtered using drug-like properties as used in our earlier works (Onawole et al., 2018, 2017; Sulaiman et al., 2019) which include having a maximum of 5 halogen atoms, five chiral centers and ten rotatable bonds; a minimum of 10 heavy (non-hydrogen) atoms and a minimum of 1 aromatic ring; and lastly should not violate not more than one of the Lipinski's rule of five (RO5) (Lipinski, 2004; Lipinski et al., 1997). After the filtration, 100,000 compounds were screened randomly using Autodock VINA as the molecular docking tool (Trott and Olson, 2010). The diversity selection of these 100,000 compounds ensured that the maximum similarity (S) threshold was set to 0.85. This assured that none of the resulted molecules was more similar than S, based on the Tanimoto coefficient of similarity (Bajusz et al., 2015; Cerqueira et al., 2015). The top-scored compounds were kept and used

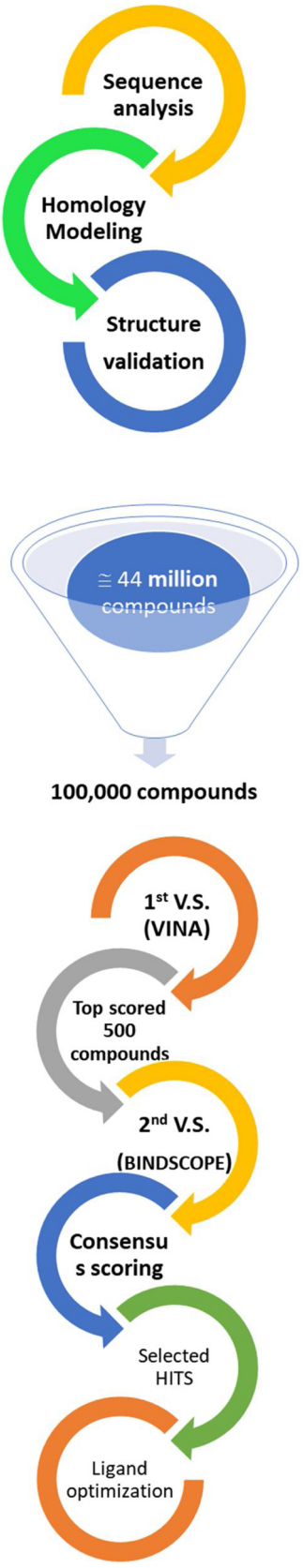

Fig. 1. Flowchart depicting the methodology.

for a second virtual screening with BindScope (Jiménez Luna et al., 2018; Mysinger et al., 2012; Skalic et al., 2018). The second virtual screening, BindScope, employed a machine learning technique (convoluted neural network). Recently, machine learning technique is gaining considerable attention in the drug discovery process as it is much faster than molecular docking. Moreover, having a different method rather than another molecular docking program with a different scoring method helps to reduce the number of false positives (Chen et al., 2018; Stevens, 2014). The latest iteration of DUD-E database (Mysinger et al., 2012) was used in the training of BindScope. This database was comprised of 22,886 active compounds and 50 similar decoys for each active and docked against 102 different targets (Koes et al., 2013). To ensure a fair benchmarking, the targets were clustered by employing a 70 % sequence similarity cut-off which was provided by blastclust in the RSCB PDB database (www.rscb.org, Burley et al., 2018). To use BindScope, the target protein, the homology modeled structure of the spike protein of SARS-CoV-2 in this case, was uploaded in PDB to the web application alongside a set of docked ligands (in this case the top 500 scored ligands from the first virtual screening) in structure-data file (SDF) format. The results from the virtual screenings were combined for consensus scoring using the rank voting method (Feher, 2006) to select the compounds which appeared as top-scored in both virtual screenings. The consensus scoring method is known to improve HIT rates by reducing the chances of false positives (Charifson et al., 1999; Huang et al., 2010; Stevens, 2014; Yang and Hsu, 2005) and has been applied in finding potential inhibitors of protein kinase B in anti-cancer drug discovery (Forino et al., 2005). The Discovery studio program (BIOVIA, 2015) was used to visualize the molecular interactions of the selected HITS with that target protein. Fig. 1 shows the flowchart that depicts the entire process of selecting hit compounds.

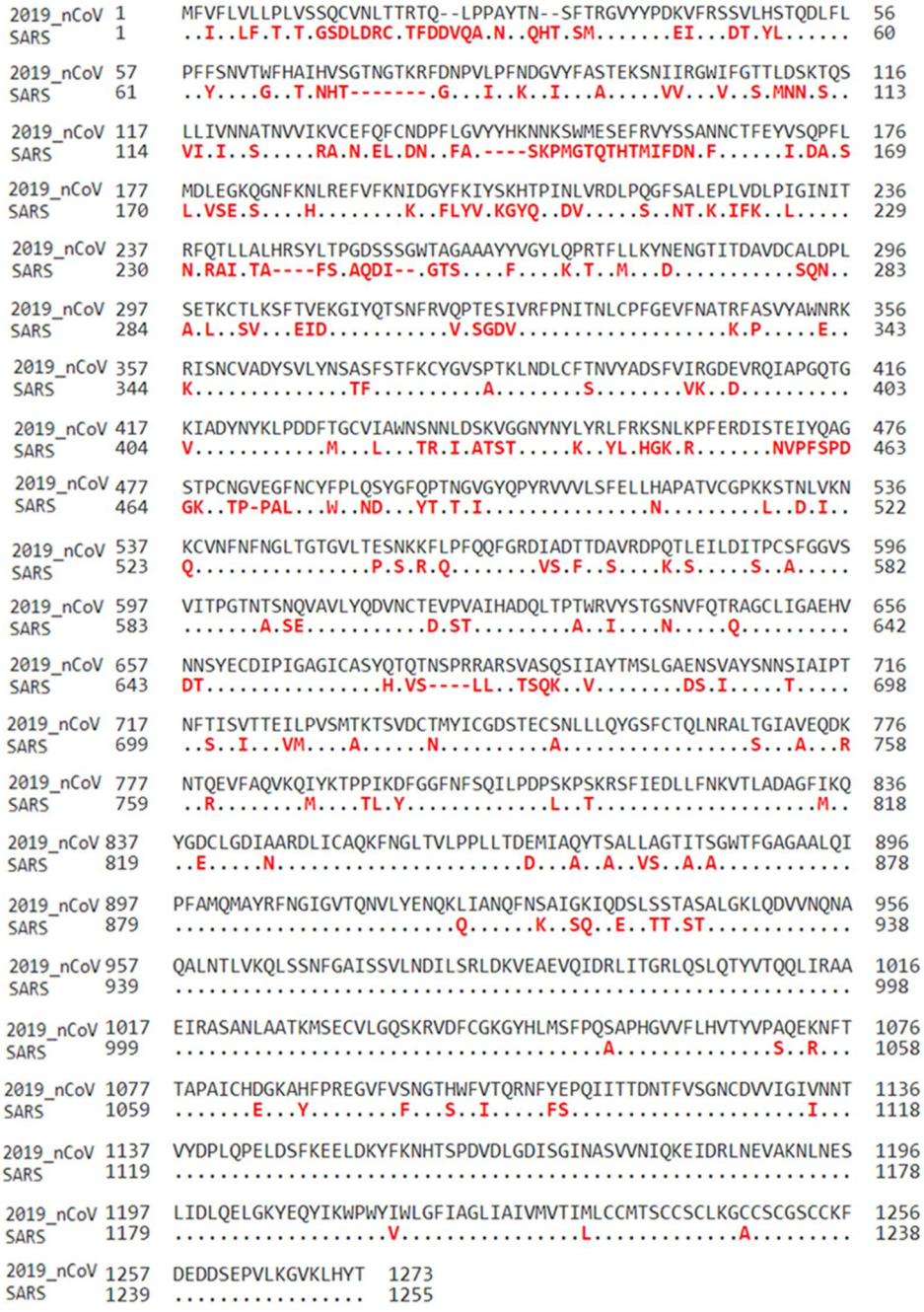

Fig. 2. Sequence alignment of the surface glycoprotein for both SARS-CoV-2 and SARS-CoV corona viruses. Identical residues are denoted by an "." beneath the consensus position.

3. Results and discussion

3.1. Sequence alignment analysis

From the whole genome of the SARS-CoV-2, only the genomic sequence of the surface glycoprotein was considered as this is the critical part of the virus responsible for virus entry and binding (Gralinski and Menachery, 2020), followed by a multiple sequence analysis using the Needleman-Wunch pairwise alignment method (Altschul et al., 2005, 1997) to compare the surface glycoprotein of both the SARS-CoV-2 and the SARS-CoV. The surface glycoprotein for SARS-CoV-2 is made up of 1273 amino acids while the SARS-CoV is 1255. The SARS-CoV-2 is 76 % identical to the SARS-CoV (Fig. 2). That is, 970 amino acids present in SAR-CoV-2 are equally present in SARS-CoV. The dots depicted in the multiple sequence alignment (MSA) for SARS-CoV denote identical sequences with 2019-nCoV (i.e. SARS-CoV-2) while the differences are denoted in red (Fig. 2). The percentage similarity suggests how closely related these two viruses are and also gives insights into how successful approach used in curtailing the SARS-CoV may also be effective for the novel SARS-CoV-2.

Besides, further MSA was done to compare the spike protein of SARS-CoV-2 with other related surface glycoproteins (SARS-CoV) with the aid of the ConSurf server (Ashkenazy et al., 2016). The phylogenetic tree which shows the evolutionary connection between the SARS-CoV-2 and other related species (Fig. 3) denote that the spike protein of SARS-CoV-2 is closely related to other coronaviruses such as bat SARS and middle east respiratory syndrome (MERS) corona virus (Table S1) but it is closest to UniRef90 A0A2R3SUW7 (The UniProt Consortium, 2019) which is a Bat SARS-like coronavirus. The phylogentic tree suggests that SARS-CoV-2 may have originated from bats (Fan et al., 2019). The MSA of SARS-CoV-2 was done with the other 29 species depicted in the phylogentic tree (Table S8). The conservation score which is a score allotted to each amino acid in a MSA is used to determine how conserved the amino acid is. A score value of 9 (maroon color) means the amino acid is well conserved while a value of 1 (cyan color) means it is variable (Fig. 4). The conservation scores of SARS-CoV-2 (Table S2) denote that about 60 % of the amino acids are conserved, that is, they have a conservation score of at least 6.

3.2. Protein structure and validation

The tertiary structure of the surface glycoprotein shows a few α-helices, many β-pleated sheets, and long random coils (Fig. 5a). The Ramachandran plot analysis, which depicts the favored, allowed and outlier values of ψ against φ angles for a particular amino acid (Ramachandran and Sasisekharan, 1968), was used to validate the structure of the protein. A good quality protein is expected to have an outlier of less than 5 % (Kleywegt and Jones, 1996). For the SARS-CoV-2 surface glycoprotein, the homology modeled structure showed 93.6 %, 5.5 % and 0.9 % in the favored, allowed and outlier regions respectively. The outlier region which is less than 5% validates the choice of protein structure for virtual screening analysis. The deeper and lighter shade of blue and orange depicts the favored regions and allowed regions respectively (Fig. 5b). The triangles and squares are the general/Pre-Pro/Proline amino acids whereas the crossed-x denotes the glycine amino acids. The eleven amino acids which make up the outliers are in red squares. They are all found in the General and Pre-Pro areas, and none occurred in the Glycine area. Nevertheless, the protein structure is good enough for further analysis.

3.3. Consensus scoring

The first virtual screening employed a molecular docking technique with the aid of VINA (Trott and Olson, 2010) and the top 500 scored compounds were used for a second virtual screening but this time, using a machine learning technique (CNN) in the BindScope web tool (Jiménez Luna et al., 2018). Only the top 500 compounds were considered as the rest of the 100, 000 compounds considered had negative binding scores with the target protein. This is evident in the docking scores of the last 100 compounds from the first virtual screening (Table S2). The top 25 scored ligands from the first virtual screening using VINA (Fig. 6) and the second virtual screening using BindScope (Fig. 7) were considered. This approach is known as the vote rank method in consensus scoring (Feher, 2006). The top scores in VINA correlate to the ligands with the highest binding energies, while the more negative values imply a stronger binding affinity to the target protein. Whereas, in BindScope, the top-scored ligands are based on probability where values close to 1 imply strong binding affinity and those close to 0 imply low binding affinity with the target protein. Three compounds namely MCULE-2442351665-0-1, MCULE-6855995445-0-2 and MCULE-4671321297-0-1 appear in the top 25 scored ligands for both VINA and BINDSCOPE. MCULE-2442351665-0-1 is the 15th and 13th top scored ligand in VINA and BINDSCOPE respectively while MCULE-6855995445-0-2 appears as the 24th and 17th top scored ligand in VINA and BINDSCOPE respectively. MCULE-4671321297-0-1 comes in as the 25th and 9th top scored ligand in VINA and BINDSCOPE

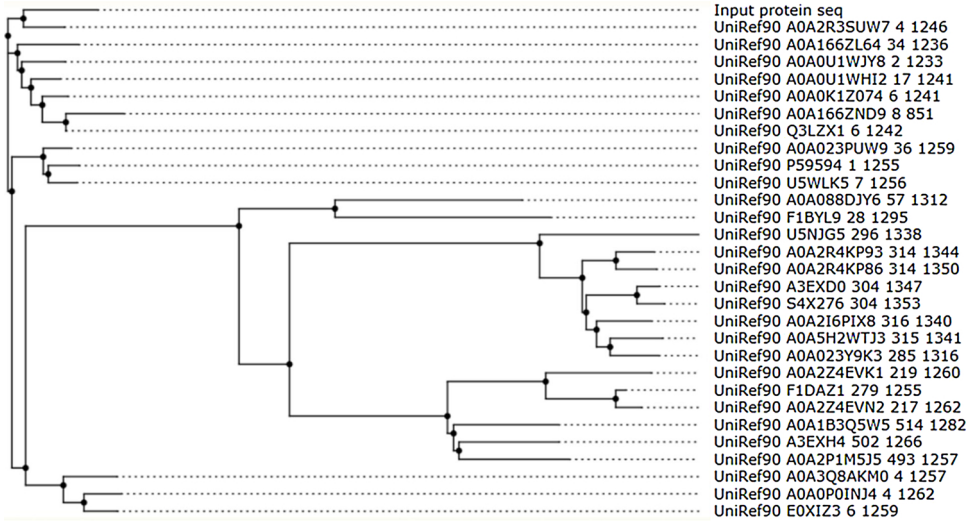

Fig. 3. The phylogenetic tree for SARS-CoV-2 (input protein sequence).

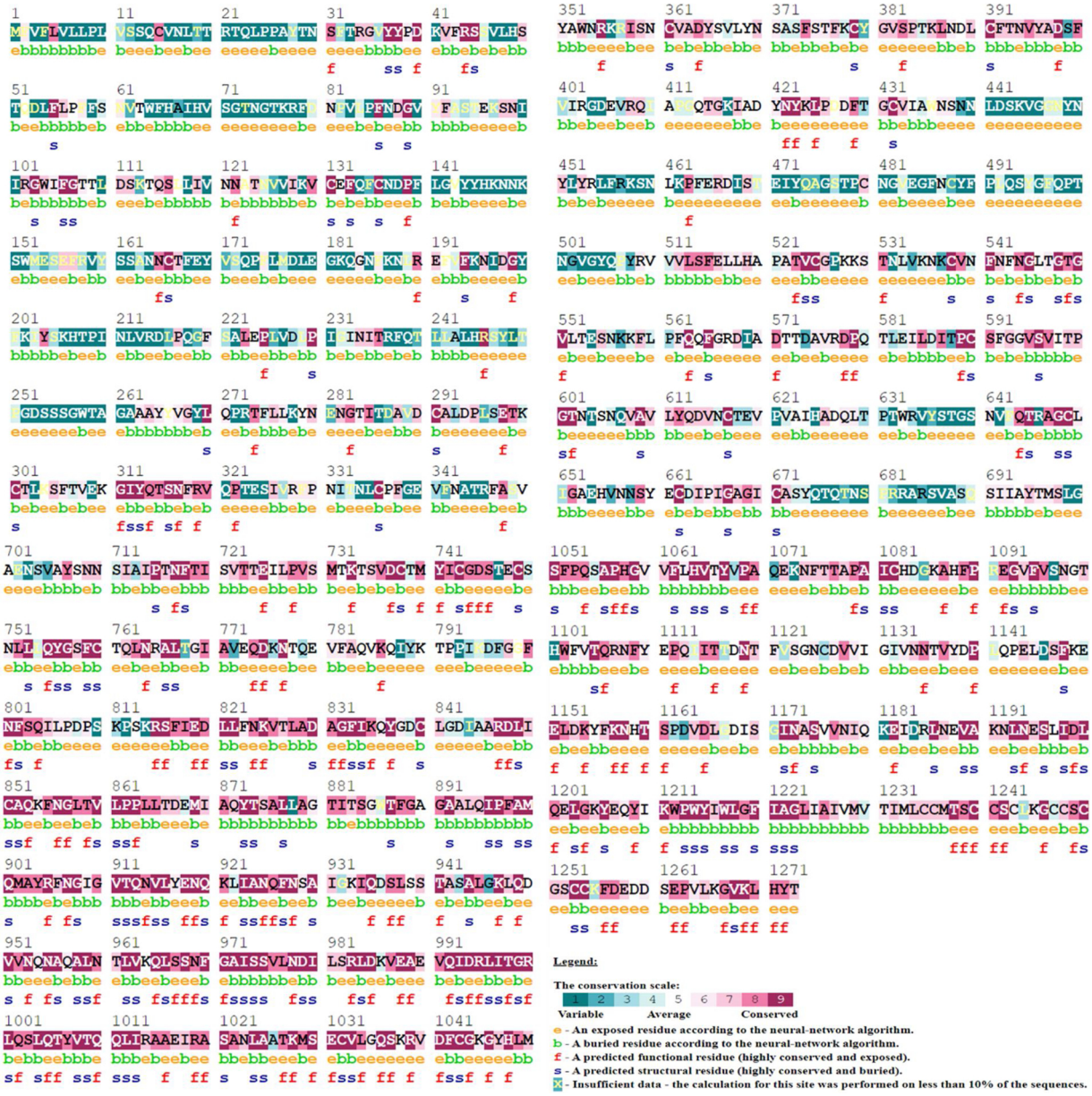

Fig. 4. The sequence the surface glycoprotein for SARS-CoV-2 showing the conservation score. The color scale represents the conservation scores where '9' implies highest conserved and '1' means highest variable.

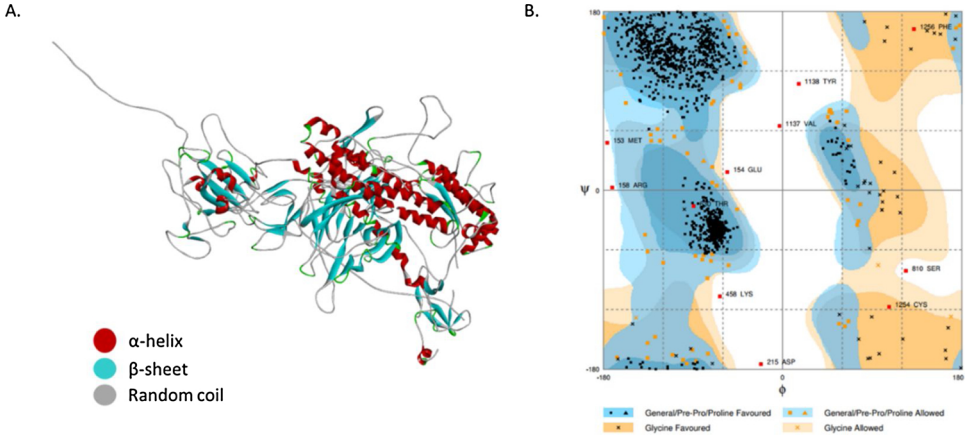

Fig. 5. The tertiary structure (A) and Ramachandran plot (B) of the surface glycoprotein of SARS-CoV-2.

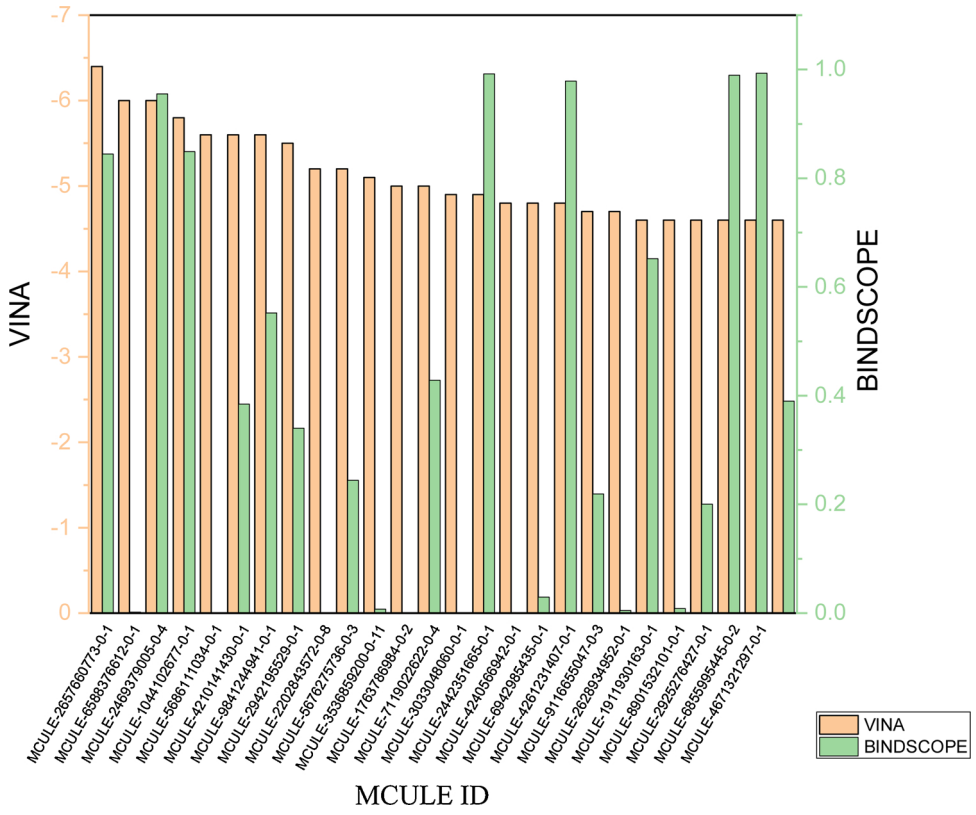

Fig. 6. The top 5% scored ligands from the first virtual screening (VINA).

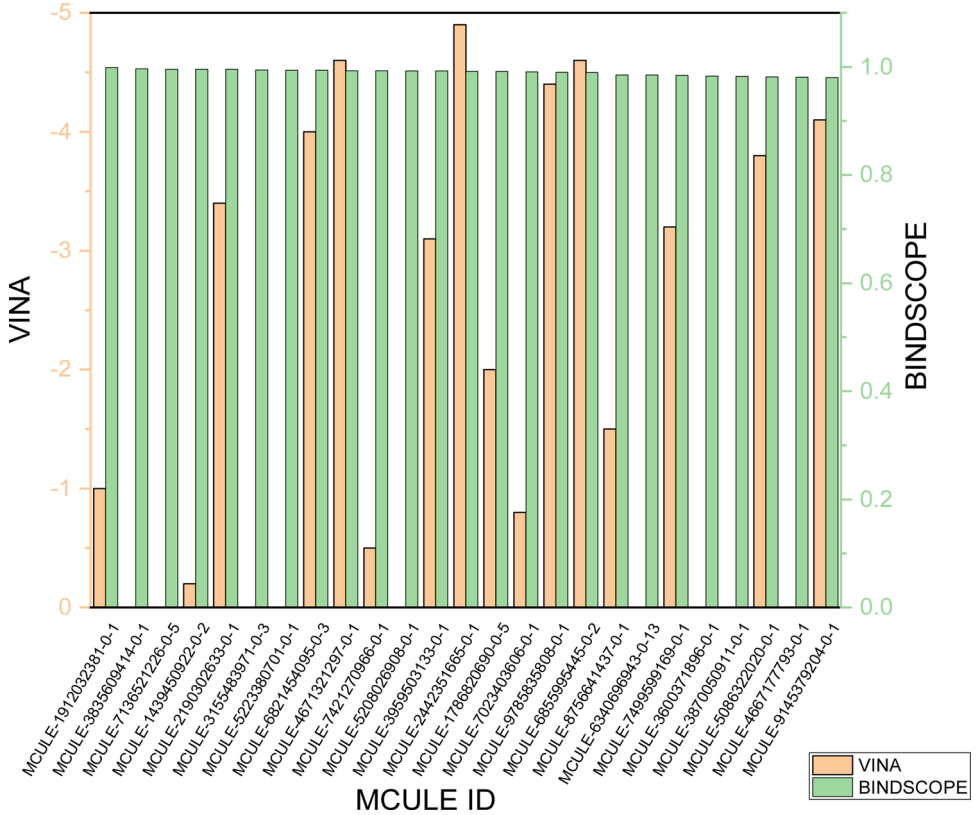

Fig. 7. The top 5% scored ligands from the second virtual screening (BindScope).

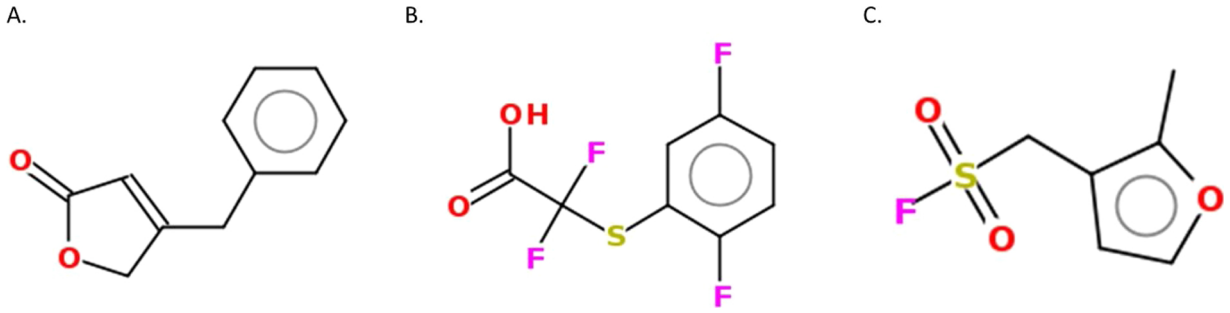

Fig. 8. The selected ligands (compounds A B and C) from consensus scoring.

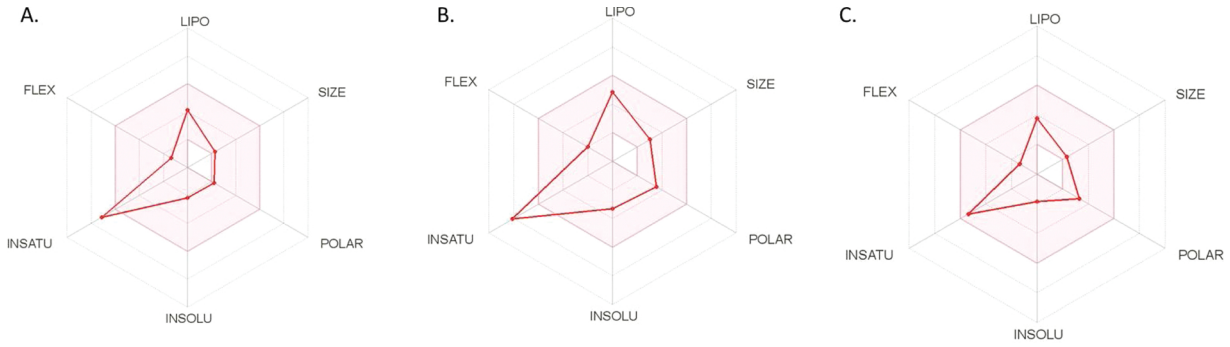

Fig. 9. The oral bioavailability radar of the selected ligands (((A) compound A, (B) compound B, and (C) compound C). The colored zone is the suitable physicochemical space for oral bioavailability.

respectively. The selected ligands, MCULE-2442351665-0-1 (benzylfuran-2(5 H)-one), MCULE-6855995445-0-2 [((2,5-difluorophenyl) thio)-2,2-difluoroacetic acid)] and MCULE-4671321297-0-1 [(2-methylfuran-3-yl)methanesulfonyl fluoride] are henceforth referred to as compounds A, B and C in the subsequent sections of this article.

3.4. Physicochemical and ADMET assessment of selected ligands

Fig. 8 presents the chemical structure of the three selected compounds. While each of the three compounds has oxygen atoms, only compound A has two rings and the other two have one ring each. All the three compounds comply with the Lipinski's RO5 (Lipinski, 2016; Lipinski et al., 2001). The oral bioavailability radar (Fig. 9) shows that the colored zone is the perfect space for the physicochemical space. The LIPO (Lipophilicity) is derived from the XLOGP3 parameter (Table 1) and is expected to be in the range of -0.7 to +5.0. The values for all the selected compounds fall within this required range and so are in the colored region. For the SIZE, it is expected not to exceed 500 $gmol^{-1}$ according to Lipinski's RO5, of which all the compounds obey. The POLAR (polarity) is determined by the Total Polarity Surface Area (TPSA) and the recommended range is between 20–130 $Å^2$, within which all the selected ligands fall. The INSOLU (insolubility) category shows that all the selected ligands are soluble as they all fall between the range of 0 and 6 for their log S (ESOL) values. Ditto for the FLEX (flexibility) which is determined by the number of rotatable bonds and is expected not to exceed nine. However, the INSATU (Insaturation) requirement which is determined by the fraction of carbon sp^3 (Csp^3) is expected to be in the range of 0.25 and 1, and this is met by compound C only. Hence, compound C has the best oral bioavailability since all its physicochemical parameters are in the colored zone.

Table 1
The physicochemical properties of the selected ligands.

Ligand	Compound A	Compound B	Compound C
Formula	$C_{11}H_{10}O_2$	$C_8H_4F_4O_2S$	$C_6H_7FO_3S$
VINA	−4.9	−4.6	−4.6
BINDSCOPE	0.9919	0.9896	0.9929
Mass	174.2	240.17	178.18
#Heavy atoms	13	15	11
#Rotatable bonds	2	3	2
#H-bond acceptors	2	6	4
#H-bond donors	0	1	0
TPSA	26.3	62.6	55.66
XLOGP3	1.69	2.97	1.1
WLOGP	1.71	4.42	2.74
ESOL Log S	−2.19	−3.3	−1.84
ESOL Class	Soluble	Soluble	Very soluble
Lipinski #violations	0	0	0
Bioavailability Score	0.55	0.56	0.55
PAINS #alerts	0	0	0
Synthetic Accessibility	2.23	2.29	2.87

Table 2 presents the results of the ADMET (absorption, distribution, metabolism, excretion, and Toxicity) analysis that was done using the AMDETSAR and SWISS ADME web tools (Cheng et al., 2012; Daina et al., 2017; Yang et al., 2018). The green-colored cells indicate excellent ADMET properties; the blue means good while the yellow and pink signify caution is needed and slightly dangerous respectively. These color codes may help during lead optimization to know what properties need to be modified. For all three selected ligands, they have good absorption properties, particularly concerning their human oral bioavailability which has been earlier suggested from the bioavailability radar. For distribution, the selected ligands are all permeants of the blood-brain barrier (BBB). Furthermore, the selected ligands are not substrate for the P-gp or the multidrug resistance protein that is responsible for transporting substances across the cell membrane. Hence, they can quickly move across the cell membrane. The metabolism of the selected compounds was predicted for cytochrome P450 inhibitors which catalyze many reactions involve in the metabolism of drugs. Compounds A and B are predicted to be inhibitors of CYP1A2 inhibitor while for the other cytochromes, all three selected ligands are non-inhibitors. The inhibition of Compounds A and B will increase plasma concentration and may lead to adverse outcomes. However, for all other cytochrome P450, the selected ligands are non-inhibitors which

Table 2
The ADMET predictions of the selected ligands: compounds A, B, and C.

ADMET Absorption	Compound A Remark (Probablity)	Compound B Remark (Probablity)	Compound C Remark (Probablity)
Human Intestinal Absorption (HIA)	Good (0.99)	Good (0.91)	Good (0.96)
Human oral bioavailability (HOB)	Good (0.76)	Safe (0.63)	Safe (0.79)
Caco-2 permeability	Good (0.91)	Good (0.60)	Good (0.65)
Distribution			
Plasma protein binding	Good (0.77)	Good (0.97)	Good (0.89)
BBB permeant	Yes	Yes	Yes
P-glycoprotein (P-gp) substrate	No	No	No
Metabolism			
Cytochrome (CYP450)			
CYP1A2 inhibitor	Yes	Yes	No
CYP2C19 inhibitor	No	No	No
CYP2C9 inhibitor	No	No	No
CYP2D6 inhibitor	No	No	No
CYP3A4 inhibitor	No	No	No
Excretion			
No info available	–	–	–
Toxicity			
Organ Toxicity			
Human ether-a-go-go inhibition	Safe (0.71)	Safe (0.73)	Safe (0.79)
Acute Oral Toxicity class III	Slightly toxic (0.74)	Slightly toxic (0.60)	Slightly toxic (0.42)
Eye irritation	Irritating (0.97)	Irritating (0.86)	Irritating (0.95)
Genomic Toxicity			
Carcinogenicity	Safe (0.80)	Safe (0.70)	Safe (0.61)
Ames mutagenesis	Caution (0.51)	Safe (0.73)	Safe (0.64)
Eco-Toxicity			
Biodegradation	Safe (0.78)	Safe (0.85)	Safe (0.55)

will aid their metabolism as potential drugs. The human ether a-go-go (hERG) inhibition is related to ventricular arrhythmia, and it can be fatal if a drug is an inhibitor (Sanguinetti and Tristani-firouzi, 2006). The selected compounds are all non-inhibitors of hERG with compound C having the highest probability of not being a hERG inhibitor. However, they all have acute oral toxicity of class three which implies that they are predicted to be slightly toxic and irritating. This irritation is further proposed as they are all eye irritants. Nevertheless, they are all non-carcinogenic with compound A having the highest probability of not being a carcinogen. Though, compound A is also the only one among the selected compounds predicted to be mutagenic. Concerning ecological toxicity, all the selected compounds are biodegradable.

3.5. Binding modes and molecular interactions

The binding mode and molecular interaction give insight into the mode of action of the selected ligands to treating SARS-CoV-2. Surprisingly, both the binding mode and the molecular interactions for both VINA and BINDSCOPE are the same for the selected ligands. This is evident in the PDB structures (supporting files attached). The binding mode (Fig. 10) shows the orientation of the ligands in 3-D and the amino residues in the binding pocket. However, the 2-D diagram of the molecular interactions (Fig. 11) gives more details into the possible mode of action. Compound A has four favorable interactions which include a conventional hydrogen bond with THR 63, a π-alkyl bond with PRO 85, a π-donor hydrogen bond with TRY 269, and a π-π T-shaped interaction with PHE 592. Unlike compound A, compound B has one unfavorable interaction with ARG 102 and four favorable interactions. The four favorable interactions for compound B includes one π-σ bond with ILE 119, and three π-alkyl bonds with ILE 203, VAL 227 and ILE 1013. Compound C has six favorable interactions; the highest amongst the three selected ligands. These interactions include: three π-π T-shaped bonds with ILE 119, ILE 128, and ILE 203; two π-alkyl bonds with VAL 227 and ILE 1013; and one π-sulfur bond with TRP 104. The ranking of the ligands according to BINDSCOPE (Fig. 7), may be based on the number of favorable interactions as compound C which has the highest number of favorable interactions, has the highest top-scored, whereas compound B which has one unfavorable interaction is the least scored amongst the three. However, compound A is the only one that has the conventional hydrogen bond interaction which is also the shortest bond length (3.34 Å) of favorable interaction. Hence, it may be the reason it has the highest binding energy according to VINA ranking (Fig. 6) amongst the three selected ligands.

The conservation scores (Table S2) of the amino acid residues

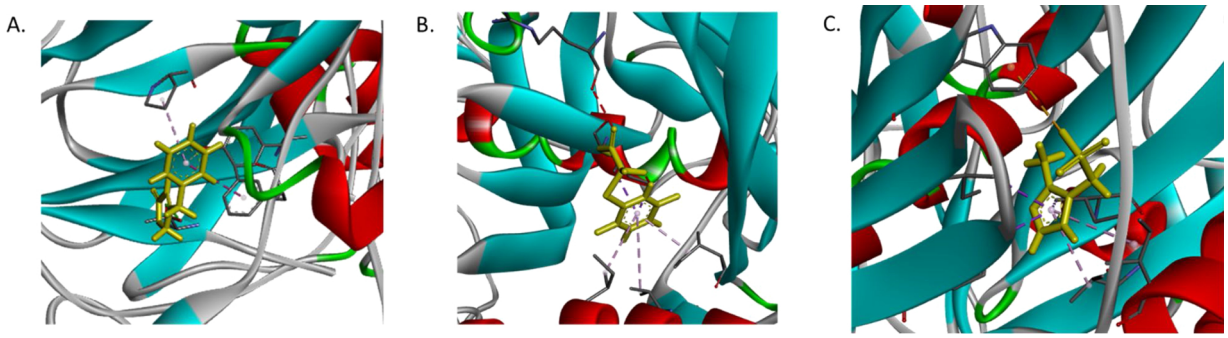

Fig. 10. The 3-D binding modes of the selected ligands ((A) compound A, (B) compound B, and (C) compound C) respectively and their molecular interactions (dashed lines) with the amino residues present in the binding pocket of the spike protein of SARS-CoV-2. The selected ligands are highlighted in yellow.

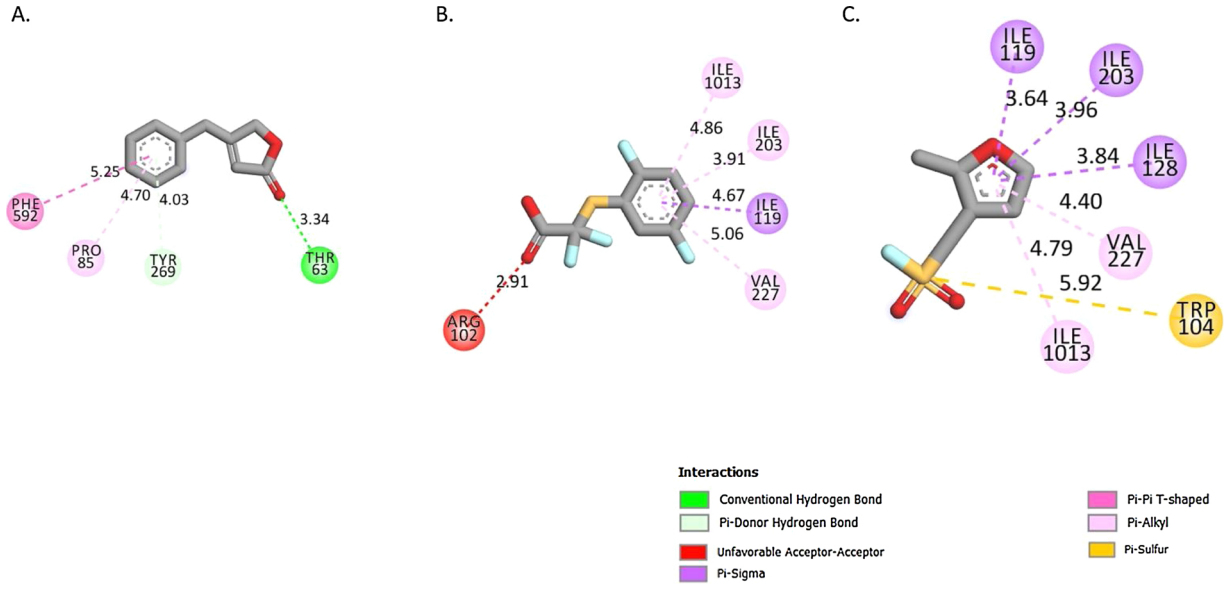

Fig. 11. The 2-D molecular interactions of the selected ligands ((A) compound A, (B) compound B, and (C) compound C) with the amino residues presnt in the in the binding pocket of the spike protein of SARS-CoV-2. The bond distances are in Angstrom (Å).

involved in the ligand-protein interactions with the selected compounds show that ILE 119, ILE 128, TRP 104 have conservation score of 7 and ILE 1013 has a conservation score of 9. This implies that the amino residues that make specific interactions with the selected compounds are conserved. The PrankWeb tool (Jendele et al., 2019) for binding pocket prediction was also used to validate the amino residue involved in the ligand-protein interaction by comparing the amino residues in the binding pocket of the crystal structure of the spike protein deposited in the protein databank (PDB ID: 6VSB). The amino residues ARG 102 and ARG104 (Fig. 11) which are involved in the ligand-protein interaction are also predicted to be in the binding pockets of 6VSB.

Moreover, to show reliability of the homology model in this study, Compound A was further docked using AutoDock VINA to a monomer of the experimental crystal structure of the spike protein (PDB ID: 6VSB). The docked ligand of compound A with the experimental structure was then compared to the homology modeled structure (Fig.12). The binding sites are similar as they both occur amidst β-pleated sheets and a few α-helix coil. The 2-D molecular interaction reveals that in the experimental structure, compound A forms a π-π interaction with PHE 592 which is similar to what is observed in the homology modeled structure. However, the differences between the experimental and homology modeled structure is the shorter distance of the interaction with PHE 592 which occurs in the former. This is also responsible for the slightly higher docking score of -5.4 kcal/mol

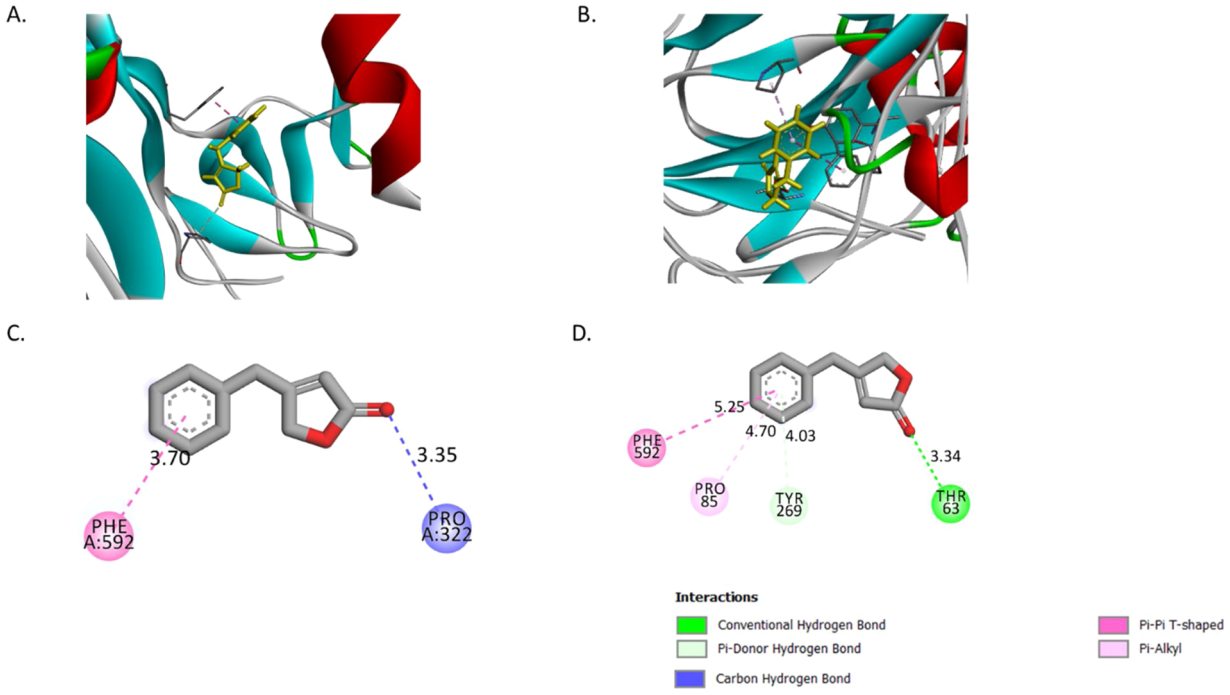

Fig. 12. The binding site of compound A in (A) experimental structure (PDB ID: 6VSB) and (B) homology-modeled structure and 2-D molecular interaction of compound A with the amino residues present in the binding site of (C) experimental structure (PDB ID: 6VSB) and (D) homology-modeled structure The bond distances are in Angstrom (Å).

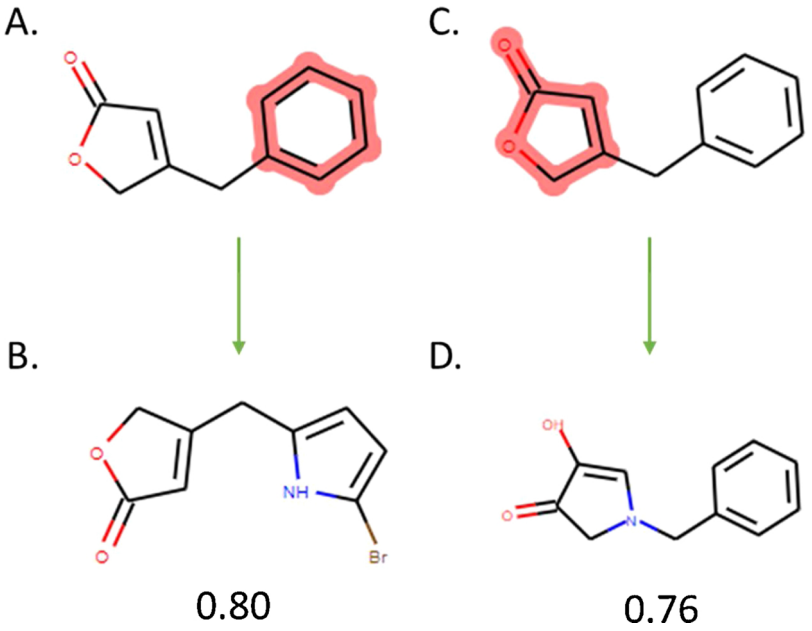

Fig. 13. The scaffold 1 (A) and scaffold 2 (C) of compound A, and the new structure after lead optimization of scaffold 1 (B) and scaffold 2 (D) respectively. The values below the new structures (B and D) are the drug-likeness values.

observed in the experimental structure. This difference in their molecular interactions is expected as the ligand would have different orientations in both structures. However, the binding of compound A to PHE 592 in both structures validates the homology-model.

3.6. Hit-to-Lead optimization

The optimization process often leads to an increase in the binding energy with the target protein, and/or an improvement in the ADMET properties. This process requires structural modifications which improve the functionality of a molecule (Jorgensen, 2009; Maynard et al., 2016; Qiao et al., 2018; Stevens, 2014). The ADMETopt webtool (Yang et al., 2018) was used for the optimization of the three selected compounds based on the ADMET properties to improve the drug-likeness. Compound A has two scaffolds, 1 and 2, as highlighted in Fig. 13. Upon optimization, the best replacement for scaffold one is a pyrol ring with a bromine subsitutent attached to it (Fig. 13B)). This replacement has the highest drug-likeness score of 0.80 (Table S4). For scaffold 2, the furan ring was replaced with a pyrole-like ring attached to a hydroxyl group. This scaffold has the highest drug-like score of 0.76 (Table S5). However, among the two scaffolds, changing scaffold 1 would led to the better optimization because of its higher drug-likeness score. Compound B only has one scaffold which is the di-fluoro benzene ring. Upon undergoing optimization using ADMETopt, the best replacement to give a highest drug-like score of 0.87 (Table S6) is replacing the scaffold with a bromine substituted pyridine (Fig. 14). For compound C, the pyrole ring with a subsititued methyl is the scaffold (Fig. 15). However, this was replaced by a thiopene ring having both methyl and bromine

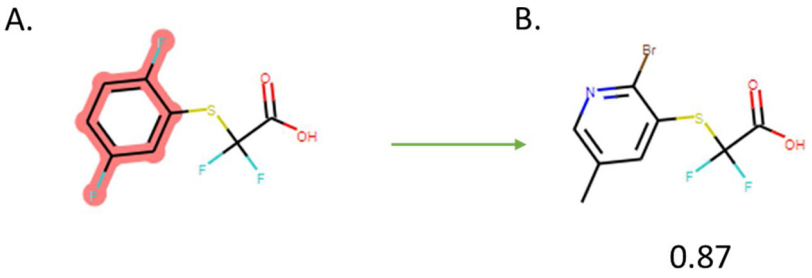

Fig. 14. (a) The scaffold (highlighted) of compound B and (b) its new structure after lead optimization. The values below the new structure denotes the drug-likeness value.

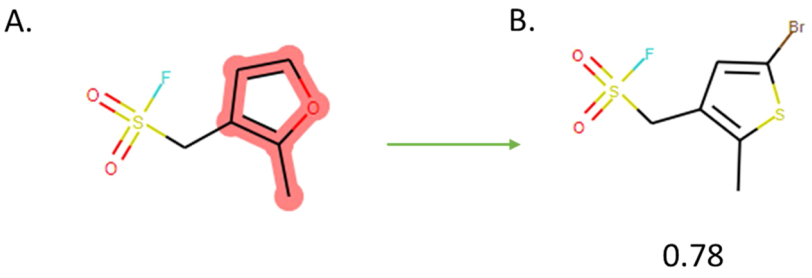

Fig. 15. (a) The scaffold (highlighted) of compound C and (b) its new structure after lead optimization. The values below the new structure denotes the drug-likeness value.

substituents attached to it has the best drug-like score of 0.78 (Table S7). In Hit-to-lead optimization, it is important to note that most times, there is usually a trade-off between improving the binding affinity or improving the drug-likeness of a molecule. In the end, the medicinal chemist needs to compromise on which area to focus on during optimization.

4. Conclusions

The recent outbreak of the COVID-19 has put all the health systems in the world on red alert as the virus spreads globally. As there are no known drugs or vaccines to treat this outbreak, developing one is paramount and computer-aided drug design is a useful tool in fast-tracking the discovery and development of new drugs that can be used to treat this disease. The consensus scoring approach has been used to combine virtual screening results from both molecular docking and machine learning to select three compounds. These compounds have the potential to inhibit the SARS-CoV-2 glycoprotein which is responsible for virus entry and binding. The molecular docking (VINA) scores of the selected compounds A, B and C are -4.9 kcal/mol, -4.6 kcal/mol and -4.6 kcal/mol, while their corresponding scores from machine learning (BINDSCOPE) are 0.992, 0.989 and 0.993 respectively. Both compounds B and C interact with amino acid residues which are conserved with ILE 1013 which is well conserved in surface glycoprotein. Compound C which has the highest score based on BINDSCOPE also has the best oral bioavailability, has all its parameters are within the recommended range. Whereas the ADMET prediction shows that the selected compounds have good absorption and distribution properties and are not carcinogenic. However, their toxicity has to be improved particularly concerning acute oral toxicity and eye irritation. These properties were considered during the Hit-to-Lead optimization which looked at the various scaffolds that can be replaced to improve the drug-likeness and non-toxicity. However, it is important to note that there is usually a trade-off between improving the binding affinity or improving the drug-likeness of a molecule. In the end, the medicinal chemist needs to compromise on which area to focus on during optimization. It is hoped that this work will help other researchers, particularly experimental medicinal scientists in developing a drug that can be used to treat the COVID-19.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Abdulmujeeb T. Onawole: Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Visualization. **Kazeem O. Sulaiman:** Conceptualization, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Temitope U. Kolapo:** Methodology, Writing - original draft, Writing - review & editing, Visualization. **Fatimo O. Akinde:** Methodology, Writing - original draft. **Rukayat O. Adegoke:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to acknowledge all the scientists, researchers, and health workers working on finding an immediate solution to treat the COVID-19, most especially those that worked to make the full genome of the virus available in a short time. Their effort is quite astonishing. They have provided the bedrock for other researchers to build on.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2020.198022.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. https://doi.org/10.1093/nar/25.17.3389.

Altschul, S.F., Wootton, J.C., Gertz, E.M., Agarwala, R., Morgulis, A., Schäffer, A.A., Yu, Y.K., 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J. https://doi.org/10.1111/j.1742-4658.2005.04945.x.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350. https://doi.org/10.1093/nar/gkw408.

Bajusz, D., Rácz, A., Héberger, K., 2015. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J. Cheminform. 7, 1–13. https://doi.org/10.1186/s13321-015-0069-3.

BIOVIA, D.S., 2015. Discovery Studio Modeling Environment.

Burley, S.K., Berman, H.M., Christie, C., Duarte, J.M., Feng, Z., Westbrook, J., Young, J., Zardecki, C., 2018. RCSB Protein Data Bank: sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. Protein Sci. 27, 316–330. https://doi.org/10.1002/pro.3331.

Cerqueira, N.M.F.S.A., Gesto, D., Oliveira, E.F., Santos-Martins, D., Brás, N., Sousa, S.F., Fernandes, P.A., Ramos, M.J., 2015. Receptor-based virtual screening protocol for drug discovery. Arch. Biochem. Biophys. 582, 56–67. https://doi.org/10.1016/j.abb.2015.05.011.

Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P., 1999. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J. Med. Chem. 42, 5100–5109. https://doi.org/10.1021/jm990352k.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., 2018. The rise of deep learning in drug discovery. Drug Discov. Today 23, 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039. PM - 29366762 M4 - Citavi.

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., Lee, P.W., Tang, Y., 2012. AdmetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. J. Chem. Inf. Model. 52, 3099–3105. https://doi.org/10.1021/ci300367a.

Clark, D.E., 2008. What has virtual screening ever done for drug discovery? Expert Opin. Drug Discov. 3, 841–851. https://doi.org/10.1517/17460441.3.8.841.

Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F., Matthew, J.B., 2002. Consensus scoring for ligand/protein interactions. J. Mol. Graph. Model. 20, 281–295. https://doi.org/10.1016/S1093-3263(01)00125-5.

Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-018-0118-9.

Daina, A., Michielin, O., Zoete, V., 2017. SwissADME : a free web tool to evaluate pharmacokinetics, drug- likeness and medicinal chemistry friendliness of small molecules. Nat. Publ. Gr. 1–13. https://doi.org/10.1038/srep42717.

Drexler, J.F., Corman, V.M., Drosten, C., 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res. 101, 45–56. https://doi.org/10.1016/j.antiviral.2013.10.013.

Fan, Y., Zhao, K., Shi, Z.L., Zhou, P., 2019. Bat coronaviruses in China. Viruses 11, 27–32. https://doi.org/10.3390/v11030210.

Feher, M., 2006. Consensus scoring for protein-ligand interactions. Drug Discov. Today 11, 421–428. https://doi.org/10.1016/j.drudis.2006.03.009.

Forino, M., Jung, D., Easton, J.B., Houghton, P.J., Pellecchia, M., 2005. Virtual docking approaches to protein kinase B inhibition. J. Med. Chem. 48, 2278–2281. https://doi.org/10.1021/jm048962u.

Gralinski, L.E., Menachery, V.D., 2020. Return of of the the coronavirus :2019-nCov. Viruses 12, 1–8.

Grosdidier, A., Zoete, V., Michielin, O., 2009. Blind docking of 260 protein-ligand complexes with EADock 2.0. J. Comput. Chem. 30, 2021–2030. https://doi.org/10.1002/jcc.21202.

Haddad, Y., Adam, V., Heger, Z., 2020. Ten quick tips for homology modeling of high-resolution protein 3D structures. PLoS Comput. Biol. 16, 1–19. https://doi.org/10.1371/journal.pcbi.1007449.

Huang, S.-Y., Grinter, S.Z., Zou, X., 2010. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. Phys. Chem. Chem. Phys. 12, 12899. https://doi.org/10.1039/c0cp00151a.

Jendele, L., Krivak, R., Skoda, P., Novotny, M., Hoksza, D., 2019. PrankWeb: a web server for ligand binding site prediction and visualization. Nucleic Acids Res. 47, W345–W349. https://doi.org/10.1093/nar/gkz424.

Jiménez Luna, J., Skalic, M., Martinez-Rosell, G., De Fabritiis, G., 2018. K DEEP : protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J. Chem. Inf. Model. 58, 287–296. https://doi.org/10.1021/acs.jcim.7b00650.

Jorgensen, W.L., 2009. Efficient drug lead discovery and optimization william. Acc. Chem. Res. 42, 724–733. https://doi.org/10.1007/s10822-014-9748-9.

Kapetanovic, I.M., 2008. Computer aided drug discovery and development: in silico-chemico-biological approach. Chem. Biol. Interact. 171, 165–176. https://doi.org/10.1016/j.cbi.2006.12.006.COMPUTER-AIDED.

Kiss, R., Sandor, M., Szalai, F.A., 2012. http://Mcule.com: a public web service for drug discovery. J. Cheminform. 4 (P17). https://doi.org/10.1186/1758-2946-4-S1-P17.

Kleywegt, G.J., Jones, T.A., 1996. Phi/Psi-chology. Structure 4. pp. 1395–1400. https://doi.org/10.1016/S0969-2126(96)00147-5. T4 - Ramachandran revisited M4 - Citavi.

Koes, D.R., Baumgartner, M.P., Camacho, C.J., 2013. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J. Chem. Inf. Model. 53, 1893–1904. https://doi.org/10.1021/ci300604z.

Krieger, E., Nabuurs, S.B., Vriend, G., 2003. Homology modeling. Struct. Bioinforma. 857, 507–508. https://doi.org/10.1007/978-1-61779-588-6.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 1–20. https://doi.org/10.1101/2020.02.19.956235.

Leelananda, S.P., Lindert, S., 2016. Computational methods in drug discovery. Beilstein J. Org. Chem. 12, 2694–2718. https://doi.org/10.3762/bjoc.12.267. PM - 28144341 M4 - Citavi.

Li, Y., Zhang, J., Wang, N., Li, H., Shi, Y., Guo, G., Liu, K., Zeng, H., Zou, Q., 2020. Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. bioRxiv. https://doi.org/10.1101/2020.01.28.922922. 01.28.922922.

Lipinski, C.A., 2004. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov. Today Technol. 1, 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007.

Lipinski, C.A., 2016. Rule of five in 2015 and beyond: target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. Adv. Drug Deliv. Rev. https://doi.org/10.1016/j.addr.2016.04.029.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. https://doi.org/10.1016/S0169-409X(96)00423-1.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 46, 3–26. https://doi.org/10.1016/S0169-409X(96)00423-1.

Liu, P., Chen, W., Chen, J.P., 2019. Viral metagenomics revealed sendai virus and coronavirus infection of malayan pangolins (manis javanica). Viruses 11. https://doi.org/10.3390/v11110979.

Lovell, S.C., Davis, I.W., Adrendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by C alpha geometry: phi,psi and C beta deviation. Proteins-Structure Funct. Genet. 50, 437–450. https://doi.org/10.1002/prot.10286.

Macalino, S.J.Y., Gosu, V., Hong, S., Choi, S., 2015. Role of computer-aided drug design in modern drug discovery. Arch. Pharm. Res. 38, 1686–1701. https://doi.org/10.1007/s12272-015-0640-5.

Manas, E.S., Green, D.V.S., 2017. CADD medicine : design is the potion that can cure my disease. J. Comput. Aided Mol. Des. 31, 249–253. https://doi.org/10.1007/s10822-016-0004-3.

Maynard, A.T., Roberts, C.D., Drive, M., Carolina, N., States, U., 2016. Quantifying, Visualizing, and Monitoring Lead Optimization. https://doi.org/10.1021/acs.jmedchem.5b00948.

Melo-Filho, C.C., Braga, R.C., Muratov, E.N., Franco, C.H., Moraes, C.B., Freitas-Junior, L.H., Andrade, C.H., 2019. Discovery of new potent hits against intracellular Trypanosoma cruzi by QSAR-based virtual screening. Eur. J. Med. Chem. 163, 649–659. https://doi.org/10.1016/j.ejmech.2018.11.062.

Mori, M., Schult-Dietrich, P., Szafarowicz, B., Humbert, N., Debaene, F., Sanglier-Cianferani, S., Dietrich, U., Mély, Y., Botta, M., 2012. Use of virtual screening for discovering antiretroviral compounds interacting with the HIV-1 nucleocapsid protein. Virus Res. 169, 377–387. https://doi.org/10.1016/j.virusres.2012.05.011.

Mysinger, M.M., Carchia, M., Irwin, J.J., Shoichet, B.K., 2012. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 55, 6582–6594. https://doi.org/10.1021/jm300687e.

Onawole, A.T., Sulaiman, K.O., Adegoke, R.O., Kolapo, T.U., 2017. Identification of potential inhibitors against the Zika virus using consensus scoring. J. Mol. Graph. Model. 73, 54–61. https://doi.org/10.1016/j.jmgm.2017.01.018.

Onawole, A.T., Kolapo, T.U., Sulaiman, K.O., Adegoke, R.O., 2018. Structure based virtual screening of the Ebola virus trimeric glycoprotein using consensus scoring. Comput. Biol. Chem. 72, 170–180. https://doi.org/10.1016/j.compbiolchem.2017.11.006.

Peng, J., Xu, J., 2011. Raptorx: exploiting structure information for protein alignment by statistical inference. Proteins Struct. Funct. Bioinforma. 79, 161–171. https://doi.org/10.1002/prot.23175.

Pereira, P., Furtado, L., Albuquerque, D., Santos, L.H.S., Antunes, D., Raul, E., Soriano, A., 2020. Structural insights into NS5B protein of novel equine hepaciviruses and pegiviruses complexed with polymerase inhibitors. Virus Res. 278https://doi.org/10.1016/j.virusres.2020.197867. 197867.

Qiao, J., Zhang, L., Hui, X., Lin, J., 2018. Kinetic and thermodynamic properties of liquid zinc: an ab initio molecular dynamics study. Comput. Mater. Sci. 141, 180–184. https://doi.org/10.1016/j.commatsci.2017.09.034.

Ramachandran, G.N., Sasisekharan, V., 1968. Conformation of polypeptides and proteins. Adv. Protein Chem. 23, 283–437. https://doi.org/10.1016/S0065-3233(08)60402-7.

Sanguinetti, M.C., Tristani-firouzi, M., 2006. hERG Potassium Channels and Cardiac Arrhythmia 440. pp. 463–469. https://doi.org/10.1038/nature04710.

Skalic, M., Martínez-Rosell, G., Jiménez, J., De Fabritiis, G., 2018. PlayMolecule BindScope: large scale CNN-based virtual screening on the web. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty758.

Sousou, J., 2015. Middle East Respiratory Syndrome Coronavirus: What Do We Know? J. Nurse Pract. 11, 131–134. https://doi.org/10.1016/j.nurpra.2014.09.019.

Stevens, E., 2014. Medicinal Chemistry: the Modern Drug Discovery Process. Pearson Education, Inc., Boston.

Sulaiman, K.O., Kolapo, T.U., Onawole, A.T., Islam, A., Adegoke, R.O., Badmus, S.O., 2019. Molecular dynamics and combined docking studies for the identification of Zaire Ebola Virus inhibitors. J. Biomol. Struct. Dyn. 37, 3029–3040. https://doi.org/10.1080/07391102.2018.1506362.

The UniProt Consortium, 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049.

Trott, O., Olson, A., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.AutoDock.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and Antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 1–12. https://doi.org/10.1016/j.cell.2020.02.058.

World Health Organization, 2020. Coronavirus latest: WHO officially names disease COVID-19.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation_supplemetary info. Science. https://doi.org/10.1126/science.abb2507.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Xiang, Z., 2006. Advances in homology protein structure modeling. Curr. Protein Pept. Sci. 7, 217–227. https://doi.org/10.2174/138920306777452312.

Yang, J.-M., Hsu, F.D., 2005. Consensus scoring criteria for improving enrichment in virtual screening. Emerging Information Technology Conference, 2005. IEEE. 2–4. https://doi.org/10.1021/CI050034W.

Yang, H., Sun, L., Wang, Z., Li, W., Liu, G., Tang, Y., 2018. ADMETopt: a web server for ADMET optimization in drug design via scaffold hopping. J. Chem. Inf. Model. acs.jcim.8b00532. https://doi.org/10.1021/acs.jcim.8b00532.

Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., Gao, G.F., 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat. Commun. 8. https://doi.org/10.1038/ncomms15092.

Zhong, N.S., Zheng, B.J., Li, Y.M., Poon, Xie, Z.H, Chan, K.H, Li, P.H, Tan, S.Y, Chang, Q, Xie, J.P, Liu, X.Q, Xu, J, Li, D.X, Yuen, K.Y, Peiris, Guan, Y., 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet (London, England) 362, 1353–1358.